Microstates to Macrodynamics
Final Report – October 5, 2010

Personnel

Joshua Weitz
Georgia Institute of Technology
jsweitz@gatech.edu
http://ecotheory.biology.gatech.edu

Jonathan Dushoff
McMaster University
dushoff@mcmaster.ca
http://lalashan.mcmaster.ca/theobio/public/index.php/Main_Page

Milestone 1: "To develop a theoretical model of collective decision making in bacterial viruses & test the ability of viruses to coordinate cell fate determination"

Result 1.1, Viral decision making: For many bacterial viruses, the choice of whether to kill host cells or enter a latent state depends on the multiplicity of coinfection. Coinfection introduces a variable number of phage DNA copies into a host cell which direct the host to produce phage mRNAs and proteins. The distinct phage genomes are coupled via a common pool of transcriptional regulators. When viral regulation of cell fate includes nonlinear feedback loops, this coupling can lead to dramatic changes in steady state gene expression. Hence, we suggest that deterministic decisions can be reached, e.g., lysis or latency, depending on the cellular multiplicity of infection, in agreement with molecular studies of the decision circuit within phage lambda (Weitz et al., Biophysical Journal, 2008; Nature Research Highlights, July 21, 2008). We are currently expanding this project in a number of different directions including sorting out how phages make decisions in a limited time horizon & developing mathematical proofs of the conditions necessary for a decision module to have a collective feature.

Result 1.2, Copy number effects on gene expression: Joshua Weitz and former postdoc, Dr. Yuriy Mileyko, have completed an in-depth analysis of the dynamical properties of small components of gene regulatory networks subject to copy number variation (see Figure below). The papers on this work have been published in Mileyko et al. (PNAS 2008) & SIAM J. Appl. Dynamical Systems (Mileyko and Weitz, 2010). Understanding expression dynamics of gene circuits for broad ranges of parameter space may provide insight into behavior of larger regulatory networks as well as facilitate the use of circuits as autonomous units performing specific regulatory tasks. In this paper, we consider three common gene circuits and investigate the dependence of gene expression dynamics on the circuit copy number. In particular, we perform a detailed bifurcation analysis of the circuits' corresponding nonlinear gene regulatory models restricted to protein-only dynamics. Employing a geometric approach to bifurcation theory we are able to derive closed form expressions for conditions which guarantee existence of saddle-node bifurcations caused by variation in the circuit copy number or copy number concentration. This result shows that the drastic effect of copy number variation on equilibrium behavior of gene circuits is highly robust to variation in other parameters in the circuits.

Milestone 2 "Develop improved models of viral-host interactions."

Results 2.1, Population and evolutionary dynamics of viral-host communities: Despite their ubiquity in natural environments, many central aspects of host-phage biology have not been integrated into mathematical models. We are developing a suite of models that will permit integrating host-phage population dynamics as a means to understand what phages do to bacteria in natural environments. (and vice-versa). First, we developed a phenomenological model of host-phage population dynamics where lysis rate depends on host density (Weitz & Dushoff, *Theoretical Ecology,* 2008*)*. We found that effective top-down control of a host-population by phages depends sensitively on the timing of invasion. Next, we investigated the effect of top-down control via viruses on the evolution of phytoplankton resource uptake (Menge & Weitz, JTB 2009). We suggest that nutrient co-limitation, and not minimum resource concentrations, are the result of evolutionary selection.

Results 2.2, Diversity patterns as emerging from molecular and ecological interactions: Trade-offs have been put forward as essential to the generation and maintenance of diversity. However, variation in trade-offs is often determined at the molecular level, outside the scope of conventional ecological inquiry. In this study, we propose that understanding the intracellular basis for trade-offs in microbial systems can aid in predicting and interpreting patterns of diversity. First, we show how laboratory experiments and mathematical models have unveiled the hidden intracellular mechanisms underlying trade-offs key to microbial diversity: (i) metabolic and regulatory trade-offs in bacteria and yeast; (ii) life-history trade-offs in bacterial viruses. Next, we examine recent studies of marine microbes that have taken steps toward reconciling the molecular and the ecological views of trade-offs, despite the challenges in doing so in natural settings. Finally, we suggest avenues for research where mathematical modeling, experiments and studies of natural microbial communities provide a unique opportunity to integrate studies of diversity across multiple scales (Gudelj, Weitz et al., 2010).

Result 2.3, Complex host and viral models of ocean microbes: We derive a simple prediction about euphotic zone N: P stoichiometry from a large class of models that use saturating nutrient uptake functions to characterize N and P acquisition by phytoplankton. The prediction is: At an ecological steady state, the ratio of phytoplankton N: P to inorganic N: P in the euphotic zone equals the ratio of phytoplankton maximum uptake rates of N and P. Comparisons of this prediction to empirical data in the surface oceans are at odds with the majority of data from extensive long-term sampling in the Atlantic and the Pacific oceans. This discrepancy calls into question the scope of applicability of ecosystem models that explicitly describe phytoplankton growth as a function of N and P availability. We discuss efforts to resolve this discrepancy, including the need for performing more comprehensive N and P uptake experiments and by reexamining models of nutrient uptake (Ballantyne et al. 2010). We then follow-up on this study by introducing flexible models of uptake, inspired by our analysis of viral attack on phytoplankton. This paper is currently in review (Menge et al, in review).

Milestone 3 "To develop a method to bin short metagenomic reads based on the statistical distribution of nucleotides & to develop indices of evolvability and modularity in microbial communities using metagenomic data."

Result 3.1, First-principles model of binning metagenomic reads: The development of effective environmental shotgun sequence binning methods remains an ongoing challenge in algorithmic analysis of metagenomic data. While previous methods have focused primarily on supervised learning involving extrinsic data, a first principles statistical model combined with a self-training fitting method has not yet been developed. We derive an unsupervised, maximum-likelihood formalism for clustering short sequences by their taxonomic origin on the basis of their k-mer distributions. The formalism is implemented using a Markov Chain Monte Carlo approach in a k-mer feature space. We introduce a space transformation that reduces the dimensionality of the feature space and a genomic fragment divergence measure that strongly correlates with the method's performance. Pairwise analysis of over 1000 completely sequenced genomes reveals that the vast majority of genomes have sufficient genomic fragment divergence to be amenable for binning using the present formalism. Using a high-performance implementation, the binner is able to classify fragments as short as 400 nt with accuracy over 90% in simulations of low-complexity communities of 2 to 10 species, given sufficient genomic fragment divergence. The method is available as an open source package called LikelyBin. An unsupervised binning method based on statistical signatures of short environmental sequences is a viable stand-alone binning method for low complexity samples. For medium and high complexity samples, we suggest combining the current method with other methods as part of an iterative process to enhance the resolving power of sorting reads into taxonomic and/or functional bins (Kislyuk et al., 2009).

Result 3.2, Identification of core and adaptive metabolism across microbial ecosystems: Metagenomic studies sequence DNA directly from environmental samples in order to understand the structure and function of complex microbial and viral communities. Individual, short pieces of sequenced DNA (``reads'') are classified into (putative) taxonomic or metabolic groups which are then analyzed for patterns across samples. Analysis of these read matrices is at the core of using metagenomic data to make inferences about ecosystem structure and function. We developed, tested and applied a non-negative matrix factorization (NMF) framework to analyze metagenomic read matrices. Non-negative matrix factorization is a numerical technique for approximating high-dimensional data points as positive linear combinations of positive canonical components. It is thus particularly well suited to biological interpretation of observed ecosystems as combinations of different components, and to disentangling overlapping patterns. We show that our method can robustly identify the appropriate degree and disentangle overlapping contributions using a synthetic data set. We then examine and discuss the NMF decomposition of a metabolic profile matrix extracted from 39 publicly available metagenomic samples comprising >4,000,000 sequence reads in total, and identify three canonical sample types, including one associated with coral ecosystems, and one that is associated with high-saline ecosystems. We also identify highly specific associations between pathways and canonical environments, and explore another NMF decomposition at a finer scale (Jeng et al., in review).

Milestone 4 – Searching for invariance and scaling in the structure of physical networks in biology, with an emphasis on plant organs, from roots to shoots to leaves

Result 4.1, Imaging and analysis platform for automatic phenotyping of plant root systems: The ability to non-destructively image and automatically phenotype complex root systems, like those of rice (Oryza sativa), is fundamental to identifying genes underlying root system architecture (RSA). Though root systems are central to plant fitness, identifying genes responsible for RSA remains an underexplored opportunity for crop improvement. We developed a non-destructive imaging and analysis system for automated phenotyping and trait ranking of RSA. Using this system we image rice roots from 12 genotypes. We automatically estimate RSA traits previously identified as important to plant function. In addition, we expand the suite of features examined for RSA to include traits that more comprehensively describe monocot RSA but that are difficult to measure with traditional methods. Using a set of 16 automatically phenotyped traits for 2297 images from 118 individuals, we observe (i) wide variation in phenotypes among the genotypes surveyed; (ii) greater inter-genotype variance of RSA features than variance within a genotype. RSA trait values are integrated into a computational pipeline which utilizes supervised learning methods to determine which traits best separate two genotypes, and then ranks the traits according to their contribution to each pair-wise comparison. This trait ranking step identifies candidate traits for subsequent QTL analysis and demonstrates that depth and average radius are key contributors to differences in rice RSA within our set of genotypes. Our results suggest a strong genetic component underlying rice RSA. This work enables the automatic phenotyping of RSA of individuals within mapping populations, providing an integrative framework for QTL analysis of RSA. (Iyer-Pascuzzi et al., 2010)

Result 4.2 Evaluating "universal" scaling relations in plant form and function
Theoretical models proposed to explain allometric relationships between organismal growth, form and function are typically tested by comparing a single predicted relationship of a given model to empirical data. Several prominent models, however, predict more than one empirical pattern, and comparisons among alternative models have not included the full range of model predictions. We rigorously evaluated several intraspecific scaling models of plant growth and form that differ in their underlying biophysical assumptions. We do this within a hierarchical Bayesian framework that simultaneously fits multiple scaling models to three large allometric datasets representing whole plants and leaves including 2362 specimens from 110 species. The scaling models we evaluated include: inflexible universal models derived from different biophysical assumptions (e.g., elastic similarity or fractal branching of internal networks), a flexible variation of a fractal network model that predicts covariation in scaling behavior, and a highly flexible model that is only constrained by basic algebraic relationships. Using the hierarchical Bayesian framework, we demonstrate that variation in allometric scaling exponents are inconsistent with any universal scaling model. We find that more flexible approaches that allow for biological variability at the species level outperform universal models, even when accounting for relative increases in model complexity (Price et al., 2009). A follow-up study considered the implications and impact of theoretical work in the area of allometric models in plant biology and the challenges that remain (Price et al., 2010).

Milestone 5 – Develop improved models to understand pathogen dynamics and evolution

Result 5.1, On the use of hemagglutination-inhibition for influenza surveillance: The hemagglutination-inhibition (HI) assay is the main tool used by epidemiologists to quantify antigenic differences between circulating influenza virus strains, with the goal of selecting suitable vaccine strains. However, such quantitative measures of antigenic difference were recently shown to have poor predictive accuracy with respect to influenza vaccine effectiveness (VE) in healthy adults. We re-examined those results using a more rigorous criterion for predictive accuracy -- considering only cases when the vaccine (V) and dominant (D) circulating strains are antigenically different -- and greater numbers of HI titers. We find that the Archetti -- Horsfall measure of antigenic difference, which is based on both the normalized HI titer (NHI) of D relative to antisera raised against V and the NHI of V relative to D, predicts VE very well ($R(2)=0.62$, $p=4.1x10(-3)$). In contrast, the predictive accuracies of the NHI of D relative to V alone ($R(2)=0.01$), and two other measures of antigenic difference based on the amino acid sequence of influenza virus hemagglutinin ($R(2)=0.03$ for both measures) are relatively poor. Furthermore, while VE in the elderly is generally high in cases when D and V are antigenically identical (VE=35%, S.E.=5%), in other cases VE appears to increase with the antigenic difference between D and V ($R(2)=0.90$, $p=2.5x10(-5)$). This paradoxical observation could reflect the confounding effects of prior immunity on estimates of VE in the elderly. Together, our results underscore the need for consistently accurate selection of suitable vaccine strains. We suggest directions for further studies aimed at improving vaccine-strain selection and present a large collection of HI titers that will be useful to such studies (Ndifon et al., 2009).

Result 5.2, Dynamics of indirectly transmitted infectious diseases: Many human pathogens rely on zoonotic hosts or have free-living stages. The dynamics of associated diseases cannot be fully understood without explicit consideration of the coupling between infected individuals and pathogen densities within environmental reservoirs. We developed and analyzed a family of epidemiological models whose mode of transmission is indirect, while accounting explicitly for the effect of the innate immune response. We base our model on recent work on the epidemiology of cholera, including that of Codeço (*BMC Infect. Dis*. 2001), expanding and generalizing previous findings in a number of important ways. First, unlike traditional SIR models, we explicitly account for the innate immune response. If the density of pathogens is smaller than the minimum infectious dose (MID), then a person will not become infected. Inclusion of a MID implies that pathogens can persist stably in environmental reservoirs without causing disease outbreaks. This is the first family of SIR models that we are aware of that explicitly incorporates a threshold-behavior due to the innate immune response. Next, disease outbreaks are of course possible for indirectly transmitted pathogens. Indeed, we find that fluctuations in pathogen density or in the fraction of infected individuals can lead to epidemics and/or endemics. De-stabilization of the disease-free state is due to global properties of the epidemiological model, in contrast to the usual finding of local instability for SIR models. We devised a set of novel measures to delineate the likelihood of outbreaks, and they serve as a guideline for future control methods. Our model suggests that for diseases transmitted through aquatic reservoirs the fluctuation of pathogen concentration might be the main driving force of epidemic outbreaks. Our finding indicates the importance of increased understanding of the in-reservoir dynamics of pathogens (Joh et al., 2009).

Publications (J.S. Weitz and **J. Dushoff** bolded)
18 published or in press, 5 submitted

Published

1. Ballantyne, F, Menge D, **Weitz JS.** (2010) A discrepancy between Michaelis-Menten based nutrient uptake model predictions and nitrogen to phosphorus stoichiometry in the surface ocean. *Limnology and Oceanography.* 55:997-1008
2. Mitchell, G.J., Nelson, D.C. and **Weitz, J.S.** (2010). Quantifying lytic enzymes: estimating the combined effects of chemistry, physiology and physics. *Physical Biology. 7:*046002.
3. Gudelj, I.[e], **Weitz, J.S.**[e], Meyer, J., Ferenci, T., Horner-Devine, M.C., Marx, C., Ackerman, M., and Forde, S.E.. (2010). An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. *Ecology Letters.* 13:1073-1084.
4. Mileyko, Y. and **Weitz, J.S.** (2010). Bifurcation analyis of gene regulatory network motifs subject to copy number variation. *SIAM J. on Applied Dynamical Systems.* 9: 799-826.
5. Iyer-Pascuzzi, A.[e1], Symonova, O.[e1], Mileyko, Y., Hao, Y., Belcher, H., Harer, J., **Weitz, J.S.**[e2], Benfey, P.N.[e2] (2010). Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. *Plant Physiology.* 152:1148-1157.
6. Boettiger, C., **Dushoff, J.** and **Weitz, J.S.** (2010). Fluctuation domains in adaptive evolution. *Theoretical Population Biology.* 77: 6-13.
7. Ndifon, W., Plotkin, J.B. and **Dushoff, J.** (2009). On the Accessibility of Adaptive Phenotypes of a Bacterial Metabolic Network. *PLoS Comput Biol.* 5: e1000472.
8. Ndifon W, **Dushoff J,** Levin SA. (2009). On the use of hemagglutination-inhibition for influenza surveillance: surveillance data are predictive of influenza vaccine effectiveness. *Vaccine.* 27:2447-52.
9. Kislyuk, A. Bhatnagar, S., **Dushoff, J.** and **Weitz, J.S.** (2009). Unsupervised statistical clustering of environmental shotgun sequences. *BMC Bioinformatics.* 10: 316.
10. Price, C.A., Ogle, K., White, E.P. and **Weitz, J.S.** (2009). Evaluating scaling theories in biology. *Ecology Letters.* 12: 641-651.
11. Menge, D. and **Weitz, J.S.** (2009). Dangerous nutrients: Evolution of phytoplankton resource uptake subject to virus attack. *Journal of Theoretical Biology.* 257: 104-115.
12. Joh, R.I., Wang, H., Weiss, H. and **Weitz, J.S.** (2009). Dynamics of indirectly transmitted infectious diseases with immunological threshold. *Bulletin of Mathematical Biology.* 71: 845-862.
13. Mileyko, Y., Joh, R.I and **Weitz, J.S.** (2008). Small-scale copy number variation and large-scale changes in gene expression. *Proceedings of the National Academy of Sciences USA.* 105: 16659-16664.
14. **Weitz, J.S,** Mileyko, Y., Joh, R.I., and Voit, E.O. (2008). Collective decision making in bacterial viruses. *Biophysical Journal.* 95: 2673-2680.
15. **Weitz, J.S.** and **Dushoff, J.** (2008). Alternative stable states in host-phage dynamics. *Theoretical Ecology*, 1: 13-19.
16. **Weitz, J.S.**, Benfey, P.N. and Wingreen, N. (2007). Evolution, interactions, and biological networks. *PLoS Biology* 5:e11.

In press

17. Price, C.A. and **Weitz, J.S.** Thickness-density tradeoffs drive specific leaf area scaling in angiosperm leaves. *American Journal of Botany.*

18. Price, C.A., Gillooly, J., Allen, A., **Weitz, J.S** and Niklas, K**.** The metabolic theory of ecology: prospects and challenges for plant biology. *New Phytologist.*

Submitted

19. Jeng, X., **Weitz, J.S., Dushoff, J.** Identification of core and adaptive metabolism across microbial ecosystems.
20. Price, C.A. Symonova, O., Mileyko, Y., Hilley, T. and **Weitz, J.S.** LEAF GUI: segmenting and analyzing the structure of leaf veins and areoles.
21. Kislyuk, A, Haegeman, B., Bergman, N. and **Weitz, J.S.** Genomic fluidity: an integrative view of gene diversity within microbial populations.
22. Joh, R.I. & **Weitz, J.S.** To lyse or not to lyse: transient-mediated stochastic fate determination in cells infected by bacteriophages.
23. Menge, D.M., Ballantyne, F.B., and **Weitz, J.S.** Predicting N:P stoichiometry from adaptive uptake models.